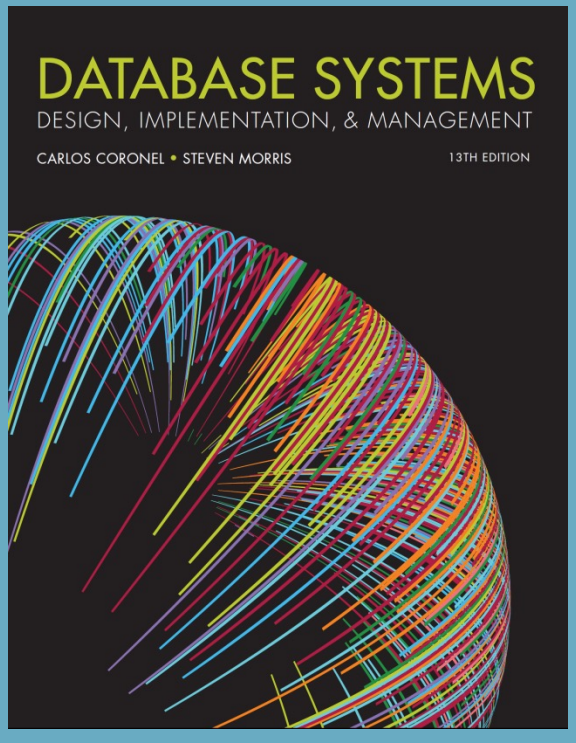


Class Intro & Chapter 1

Database Systems

Who is Professor Taylor?

- UNC BS Math Undergraduate and Johns Hopkins Masters in Computer Science

- 30 years in software development in commercial software companies

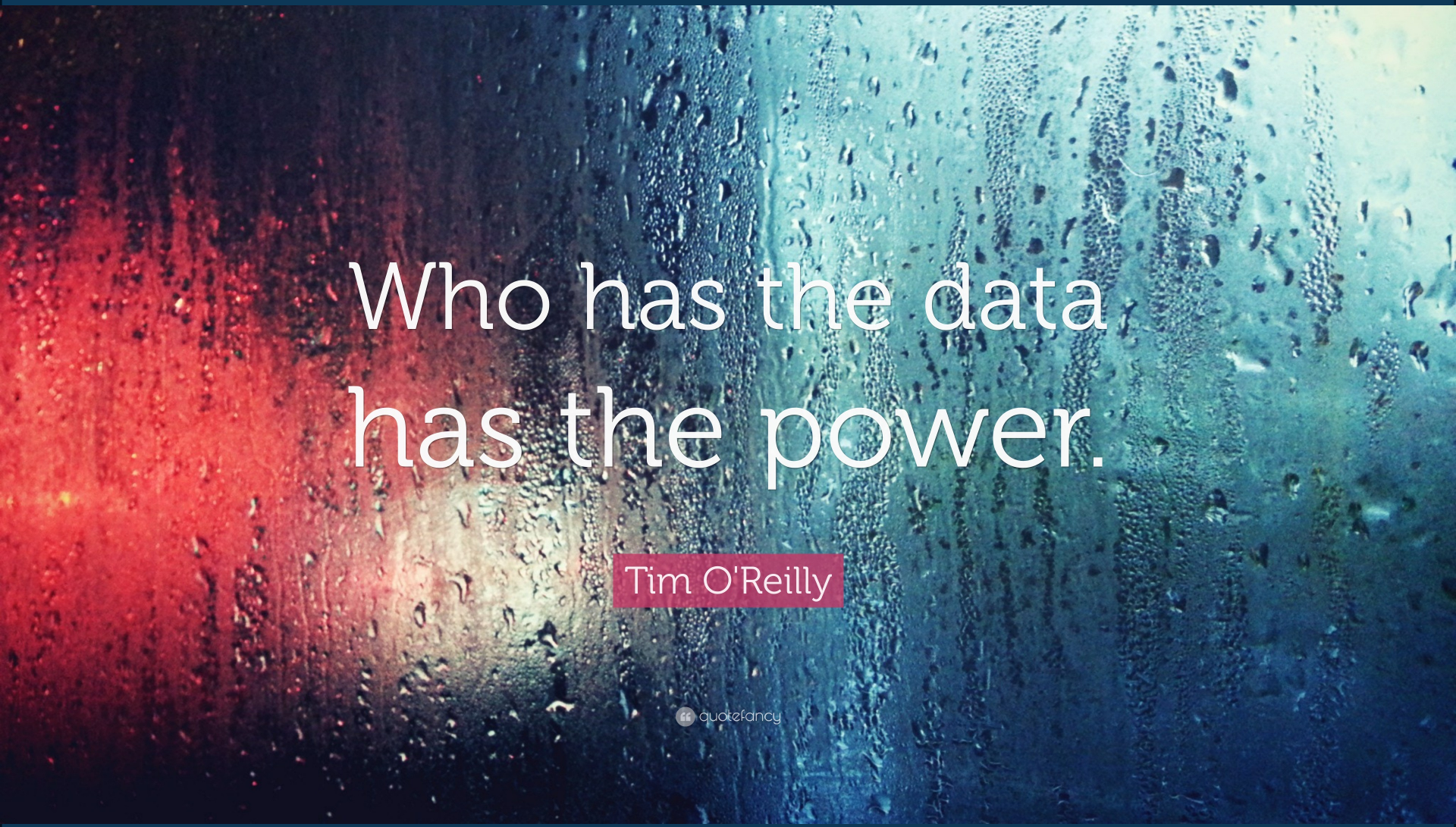
- Started as a C developer on a Top-Secret project for the Defense Department

- Managed large software teams with projects of up to 1000 team members
  - Healthcare
  - Telecom/Cable
  - eCommerce
  - Supply Chain

https://www.linkedin.com/in/constanceltaylor

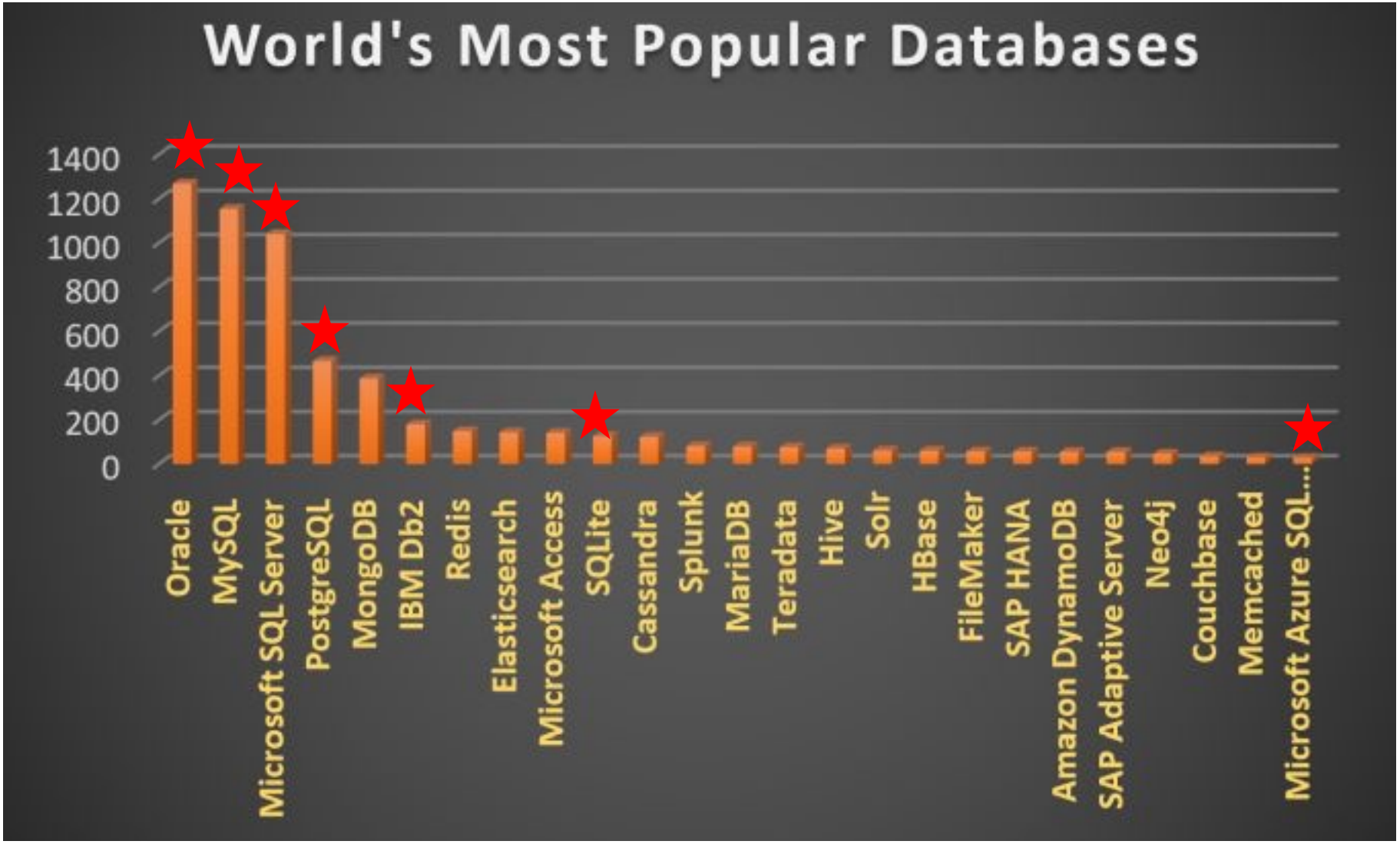
General Communication

- **Canvas**
  - Weekly announcement
- **Email**
  - connie2@clemson.edu (not Canvas or gmail)
- **Office**
  - 100H McAdams
- **Instructor Office hours:**
  - On Monday and Wednesday 1:15pm-2:15pm and virtually Tuesday from 11am-12pm. Also, by appointment.
  - When virtually connecting, I use Microsoft Teams (messaging system) – see info in Syllabus about joining Microsoft Teams.
  - On-line meetings available on request
- **TA:**
  - Abhilash Thomas
  - abhilat@g.clemson.edu
- **TA Office hours:**
  - See Canvas


Who has the data
has the power.
Tim O'Reilly
quotefancy

Course Description

- CPSC 4620&6620 will provide an introduction to database/data communications concepts as related to the design of online information systems.
- Problems involving structuring, creating, maintaining, and accessing multiple-user databases are presented and solutions developed.
- Comparison of several commercially available database management systems is made.
- This course will focus on the design and use of **relational databases and Structured Query Language (SQL).**



★ = Relational DBMS

World's Most Popular Databases - Sources

- Number of results in Google, Bing, and Yandex search engines
- Frequency of searches in Google Trends
- Frequency of technical discussions on the well-known IT-related Q&A sites Stack Overflow and DBA Stack Exchange
- The number of job offers on Indeed and Simply Hired.
- A number of profiles in professional networks, including LinkedIn and Upwork.
- Mentions in Twitter tweets.

- https://www.c-sharpcorner.com/article/what-is-the-most-popular-database-in-the-world/

Course Outline

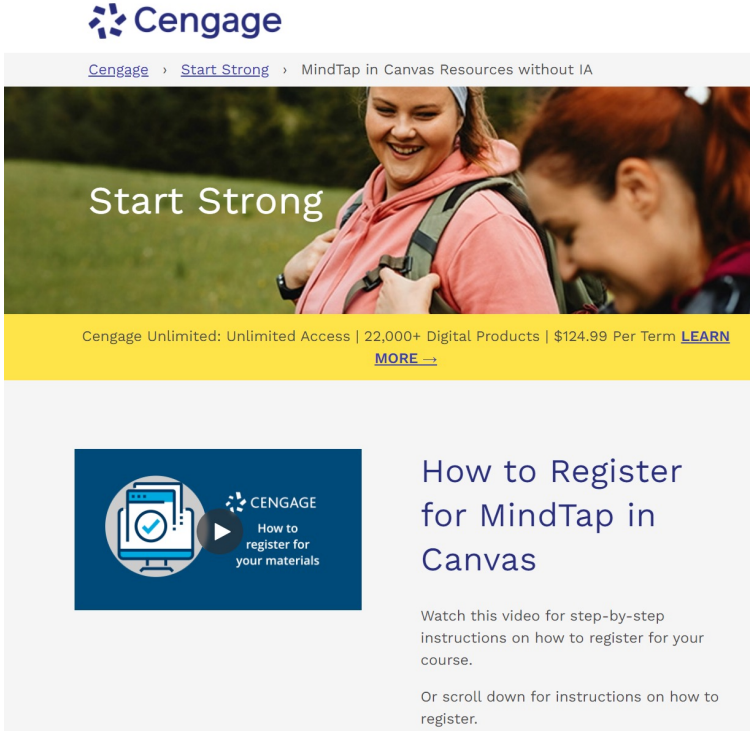
- Database Systems
- Data Models
- The Relational Database Model
- Entity Relationship (ER) Modeling
- Advanced Data Modeling
- Normalization of Database Tables
- Introduction to Structured Query Language (SQL)
- Advanced SQL
- Database Design
- Transaction Management and Concurrency Control
- Database Performance Tuning and Query Optimization
- Distributed Database Management Systems
- Business Intelligence and Data Warehouses
- Big Data and NoSQL
- Database Connectivity and Web Technologies
- Database Administration and Security

Materials

- Required text:
  - Electronic Edition + MindTap of:
    Database Systems: Design, Implementation, & Management
    by Carlos Coronel, Steven Morris
    13th Edition | Copyright 2019
    ISBN: 9780357427873
- We will cover the whole book
- **You will need to purchase this…**
  as it is an electronic textbook AND a class platform with autograding

MindTap

- Cengage MindTap Course
  - Registration for the MindTap course is done via Canvas.
  - There is a link in the first course announcement and the "Getting Started" Module in Canvas
  - Use this link for specific guidance on making the connection: https://startstrong.cengage.com/mindtap-canvas-ia-no/



CENGAGE | MINDTAP

CPSC 4620: Database Systems, 13e, V2

Expand all

- Getting Started with MindTap Video Series
- Getting Started with Coding IDE Labs Video Series
- Part 1: Database Concepts (Chapters 1 & 2)
- Part 2: Design Concepts (Chapters 3 - 6)
- Part 3: Advanced Design and Implementation (Chapters 7 - 9)
- SQL Lab
- Part 4: Advanced Database Concepts (Chapters 10 - 14)
- Part 5: Databases and the Internet (Chapter 15)
- Part 6: Database Administration (Chapter 16)

Chapter 7: Introduction to Structured Query Language (SQL)

Activity

CPSC 4620: Chapter 7 Application
COUNTS TOWARDS GRADE

CPSC 4620: Chapter 7 Review Questions
COUNTS TOWARDS GRADE

CPSC 4620: Chapter 7 Quiz
COUNTS TOWARDS GRADE

Assess your knowledge after completing Module 7.

Chapter 7 Read

Chapter 7: How it Works in MySQL

Read "How it Works in MySQL" before the Module 7 Reading if using MySQL.

Chapter 7 Reading

Read "Introduction to Structured Query Language (SQL)."

Chapter 7 Visualize

SELECT Operations

Watch engaging concept clip video to help explain and illustrate Database Systems terms and concepts.

SELECT Clause Basics

Watch engaging video tutorials to help explain and illustrate Database Systems terms and concepts.

FROM Clause Basics

Watch engaging video tutorials to help explain and illustrate Database Systems terms and concepts.

Grading

ASSIGNMENT	WEIGHT
Project (3 Parts + 1 Grad only part)	25%
Quizzes	10%
Class Exercises	10%
SQL Lab	25%
Homework	10%
Midterm and Final Exam	20%

Academic Integrity

- http://www.clemson.edu/academics/integrity/
- If you have a question if something is academically dishonest, please contact me as soon as possible

As members of the Clemson University community, we have inherited Thomas Green Clemson's vision of this institution as a "high seminary of learning." Fundamental to this vision is a mutual commitment to truthfulness, honor, and responsibility, without which we cannot earn the trust and respect of others. Furthermore, we recognize that academic dishonesty detracts from the value of a Clemson degree. Therefore, we shall not tolerate lying, cheating, or stealing in any form.

For Each Chapter

- Read the chapter
- Complete the Application Activity
  (use the link in Canvas to jump to specific assignment in MindTap)
- Complete Class Exercises in Class for credit
- Quiz in class for each chapter (uses Respondus)

SQL Lab

- Once Chapter 7 is complete there will be multiple SQL Assignments for Chapter 7 and 8
- The assignments are auto-graded in MindTap
- The assignments will span the remainder of the course

Project

- You will build a **Java** application
- The project is in 3 parts
  - design the database using an ERD
  - create the database and populate it using embedded SQL
  - complete implementation of a Java application that accesses the database (you will be provided with the interface via starter code)

Exams

- There will be a midterm and final exam
- You will need to use the Respondus Browser for exams
- The Final Exam is comprehensive and will be composed of questions from each chapter exam
- There will be a comprehensive study guide for the Midterm and Final Exam
- No exemptions from the Final Exam

DISCUSSION:

WHAT IS THE DIFFERENCE BETWEEN DATA AND INFORMATION?

WHAT IS THE DIFFERENCE BETWEEN DATA AND INFORMATION?

Data consists of raw facts

- Not yet processed to reveal meaning to the end user
- Building blocks of information

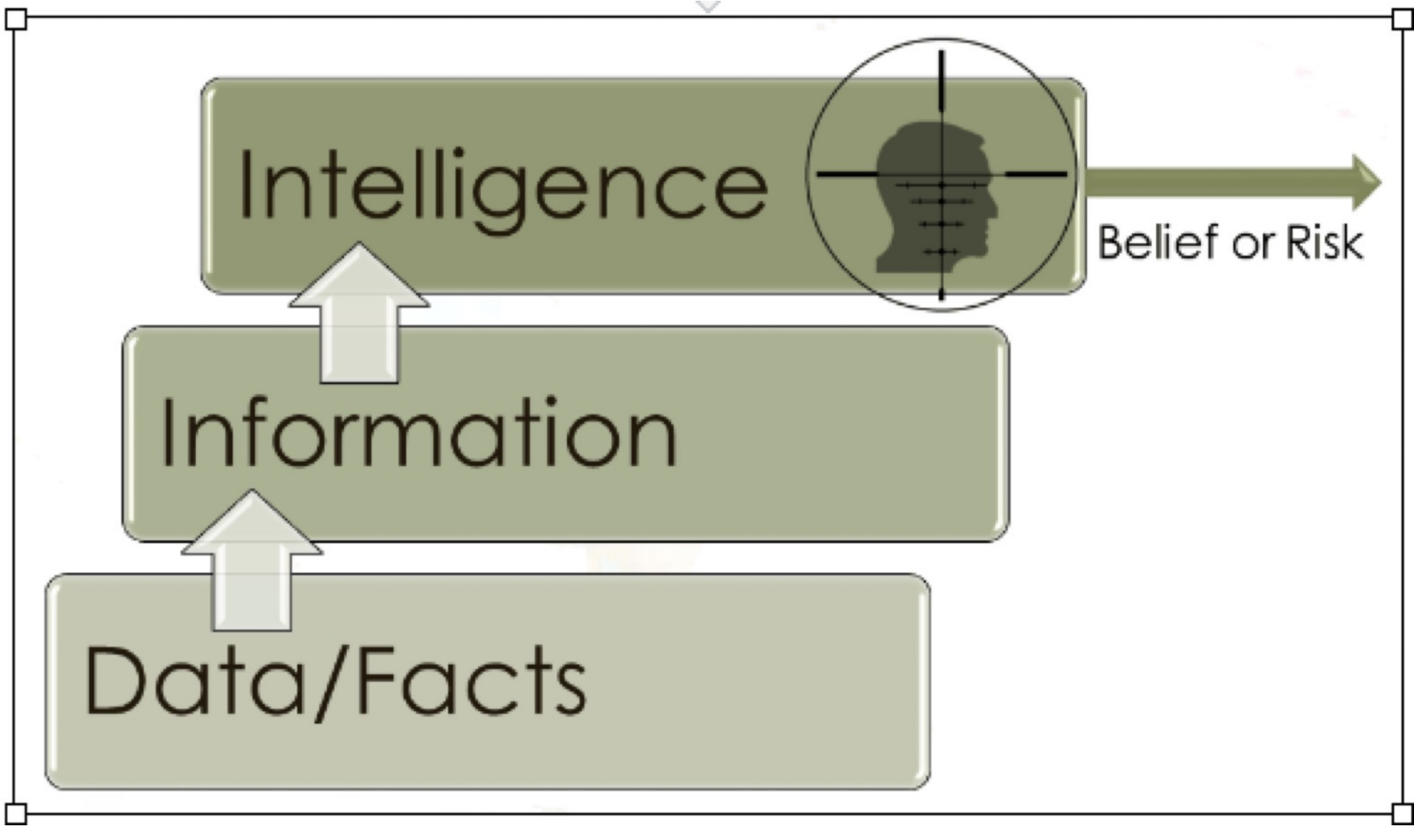
Information results from processing raw data to **reveal meaning**

- Requires context
- Bedrock of knowledge
- Should be accurate, relevant, and timely

This is done using a database…

Chapter 1 Learning Objectives

- Describe what a database is, various types, and why they are valuable assets for decision making
- Explain the importance of database design
- See how modern databases evolved from file systems
- Understand flaws in file system data management
- Outline the main components of the database system
- Describe the main functions of a database management system (DBMS)


Intelligence
Information
Data/Facts
Belief or Risk

Introducing the Database

- **Shared, integrated computer structure that stores data**
  - End-user data: raw facts of interest to end user
  - **Metadata:** data about data, through which the end-user data is integrated and managed
    - Describes data characteristics and relationships

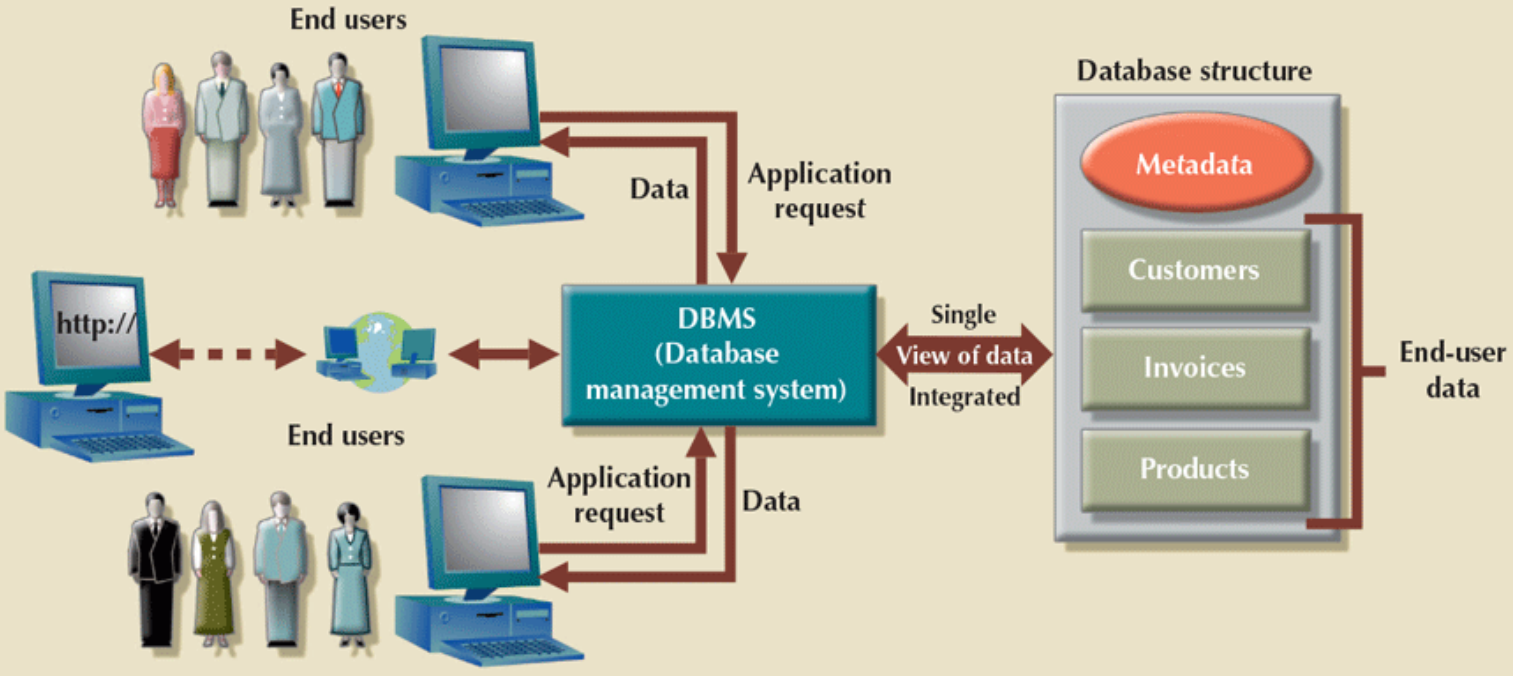
- **Database management system (DBMS)**
  - Collection of programs
  - Manages the database structure
  - Controls access to data stored in the database

Role and Advantages of the DBMS (1 of 2)

- Database management system (DBMS): intermediary between the user and the database
  - Enables data to be shared
  - Presents the end user with an integrated view of data
  - Provides more efficient and effective data management
  - Improves sharing, security, integration, access, decision-making, productivity, etc.

Role and Advantages of the DBMS (2 of 2)

FIGURE 1.4 THE DBMS MANAGES THE INTERACTION BETWEEN THE END USER AND THE DATABASE



Types of Databases (1 of 5)

There are different ways to classify databases:

- **Single-user database**: supports one user at a time
  - Desktop database: single-user database on a personal computer
- **Multiuser database:** supports multiple users at the same time
  - Workgroup databases: supports a small number of users or a specific department
  - Enterprise database: supports many users across many departments

Types of Databases (2 of 5)

- **Classification by location**
  - Centralized database: data located at a single site
  - Distributed database: data distributed across different sites
  - Cloud database: created and maintained using cloud data services that provide defined performance measures for the database

Types of Databases (3 of 5)

- Classification by data type
  - **General-purpose database:** contains a wide variety of data used in multiple disciplines
  - **Discipline-specific database:** contains data focused on specific subject areas
  - **Operational** database: designed to support a company's day-to-day operations; sometimes called online transaction processing (**OLTP**) database
  - **Analytical** database

EXAMPLES?

Types of Databases (4 of 5)

- **Analytical database**: stores historical data and business metrics used exclusively for tactical or strategic decision making
  - Data warehouse: stores data in a format optimized for decision support
  - Online analytical processing (**OLAP**): tools for retrieving, processing, and modeling data from the data warehouse
  - Business intelligence: captures and processes business data to generate information that support decision making

Types of Databases (5 of 5)

- Databases can be classified to reflect the degree to which the data is structured
  - Unstructured data exists in its original (raw) state
  - Structured data results from formatting
    - Structure is applied based on type of processing to be performed
  - Semistructured data: processed to some extent

Why Database Design Is Important

- Focuses on design of database structure that will be used to store and manage end-user data
  - Well-designed database: facilitates data management and generates accurate and valuable information
  - Poorly designed database: causes difficult-to-trace errors that may lead to poor decision making

Evolution of File System Data Processing (1 of 3)

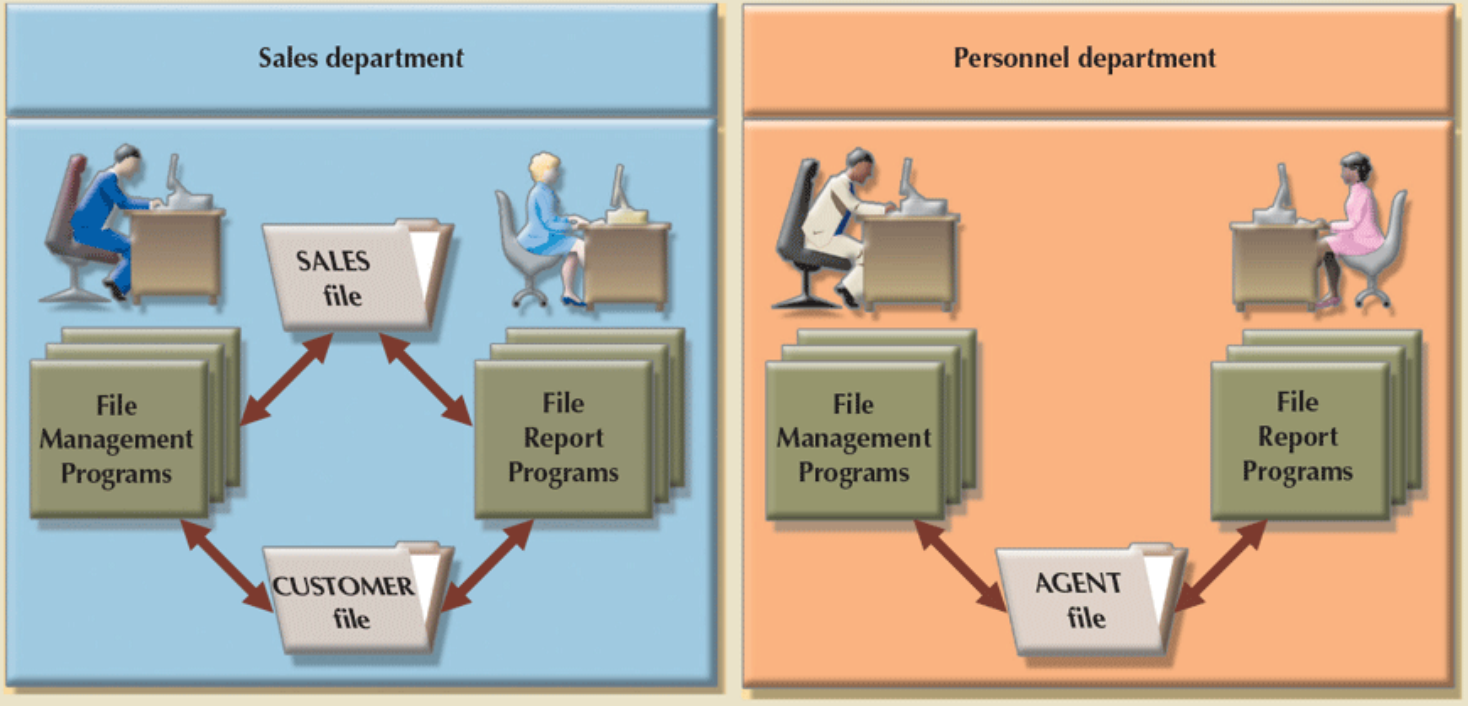
- Manual file systems
  - Accomplished through a system of file folders and filing cabinets
- Computerized file systems
  - Data processing (DP) specialist created a computer-based system to track data and produce required reports
- File system redux: modern end-user productivity tools
  - Includes spreadsheet programs such as Microsoft Excel

Evolution of File System Data Processing (2 of 3)

Table 1.2	Basic File Terminology
TERM	**DEFINITION**
Data	Raw facts, such as a telephone number, a birth date, a customer name, and a year-to-date (YTD) sales value. Data has little meaning unless it has been organized in some logical manner.
Field	A character or group of characters (alphabetic or numeric) that has a specific meaning. A field is used to define and store data.
Record	A logically connected set of one or more fields that describes a person, place, or thing. For example, the fields that constitute a record for a customer might consist of the customer's name, address, phone number, date of birth, credit limit, and unpaid balance.
File	A collection of related records. For example, a file might contain data about the students currently enrolled at Gigantic University.

Evolution of File System Data Processing (3 of 3)

FIGURE 1.9 A SIMPLE FILE SYSTEM



Problems with File System Data Processing

- File System data is a a challenge to create different types of information as well as information accuracy
  - Lengthy development times
  - Difficulty of getting quick answers
  - Complex system administration
  - Lack of security and limited data sharing
  - Extensive programming

Structural and Data Dependence (1 of 2)

- Structural dependence
  - Access to a file is dependent on its own structure
  - All file system programs are modified to conform to a new file structure

- Structural independence
  - File structure is changed without affecting the application's ability to access the data

Structural and Data Dependence (2 of 2)

- Data dependence
  - Data access changes when data storage characteristics change
- Data independence
  - Data storage characteristics are changed without affecting the program's ability to access the data
- Practical significance of data dependence is the difference between logical and physical format

Data Redundancy (1 of 2)

- Unnecessarily storing the same data at different places
  - Islands of information (i.e., scattered data locations)
  - Increases the probability of having different versions of the same data

Data Redundancy (2 of 2)

- Possible results of uncontrolled data redundancy
  - Poor data security
  - Data inconsistency
  - Data-entry errors
  - Data integrity problems

Data Anomalies

- Develop when not all of the required changes in the redundant data are made successfully
  - Update anomalies
  - Insertion anomalies
  - Deletion anomalies

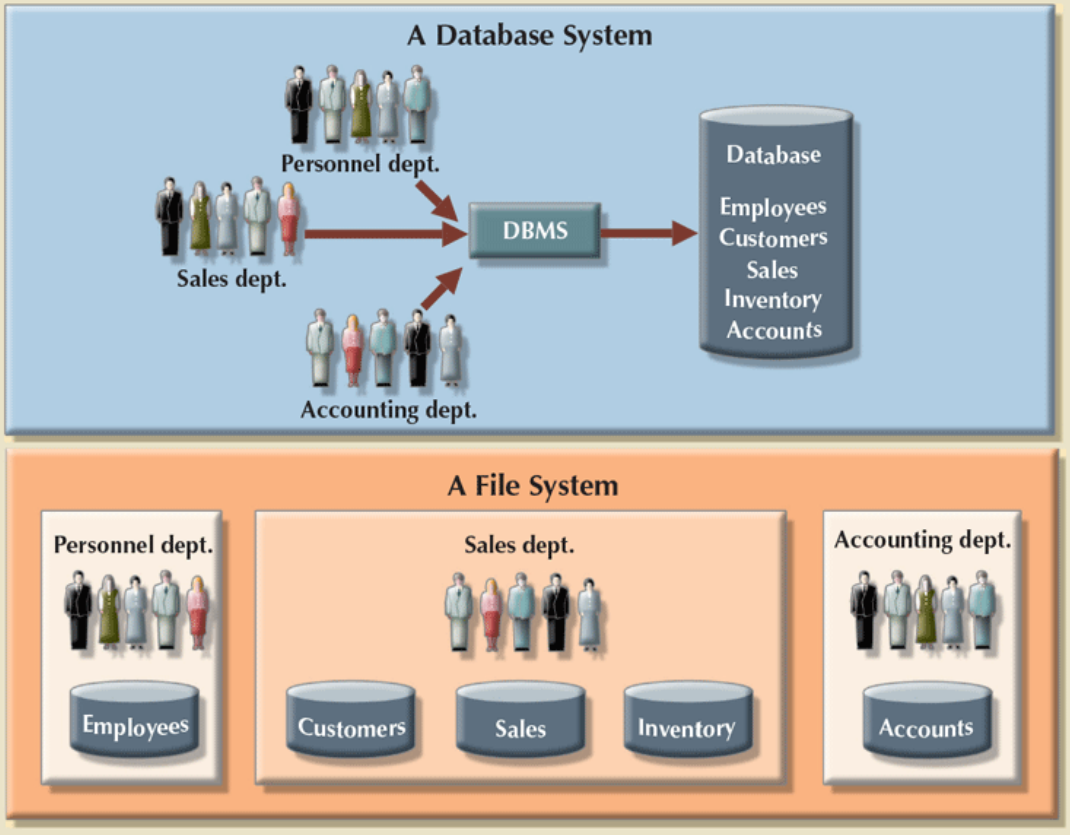
Database Systems (1 of 2)

- Logically related data stored in a single logical data repository
  - Physically distributed among multiple storage facilities
  - DBMS eliminates most of file system's data inconsistency, data anomaly, data dependence, and structural dependence problems

- Current generation DBMS software
  - Stores data structures, relationships between structures, and access paths
  - Defines, stores, and manages all access paths and components

Database Systems (2 of 2)

FIGURE 1.10 CONTRASTING DATABASE AND FILE SYSTEMS

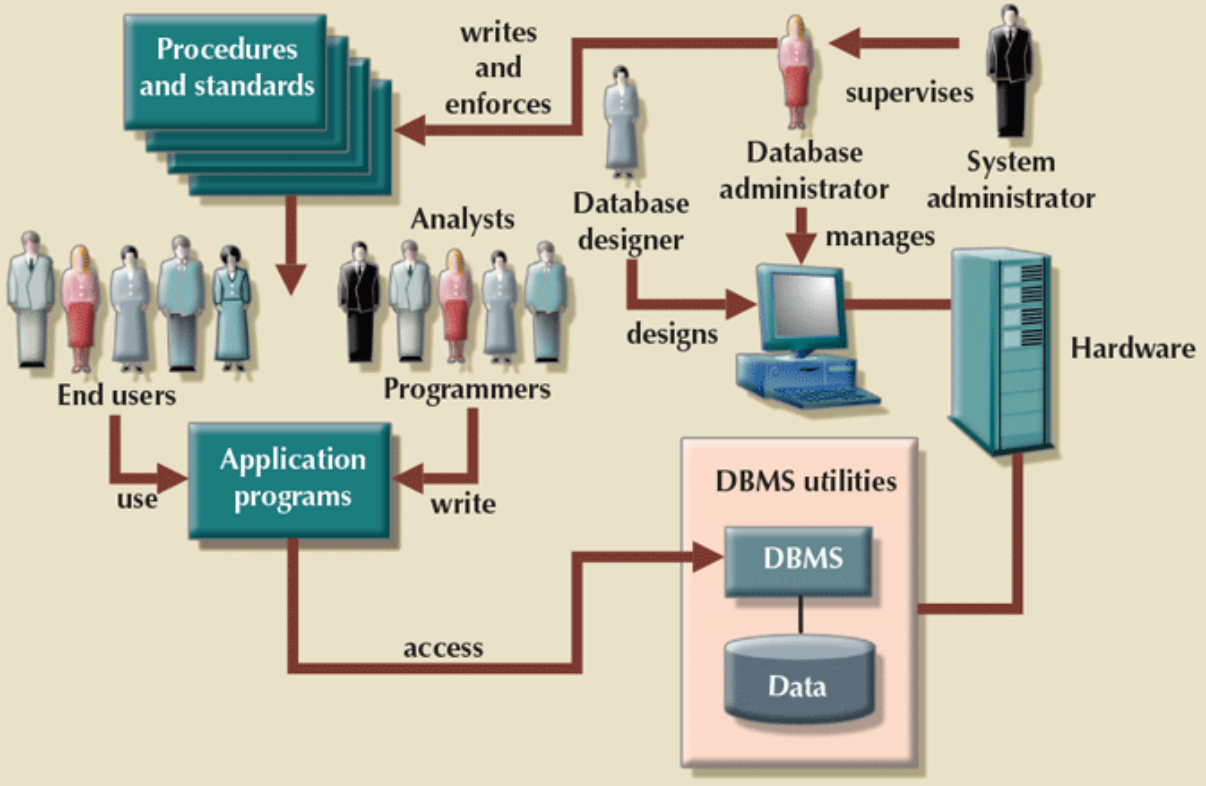


The Database System Environment (1 of 2)

- Database system: organization of components that define and regulate the collection, storage, management, and use of data within a database environment
  - Hardware
  - Software
  - People
  - Procedures
  - Data

The Database System Environment (2 of 2)

FIGURE 1.11 THE DATABASE SYSTEM ENVIRONMENT



DBMS Functions (1 of 3)

- Data dictionary management
  - Data dictionary: stores definitions of data elements and their relationships
- Data storage management
  - Performance tuning ensures efficient performance
- Data transformation and presentation
  - Data is formatted to conform to logical expectations
- Security management
  - Enforces user security and data privacy

DBMS Functions (2 of 3)

- Multiuser access control
  - Sophisticated algorithms ensure that multiple users can access the database concurrently without compromising its integrity
- Backup and recovery management
  - Enables recovery of the database after a failure
- Data integrity management
  - Minimizes redundancy and maximizes consistency

DBMS Functions (3 of 3)

- Database access languages and application programming interfaces
  - Query language: lets the user specify what must be done without having to specify how
  - Structured Query Language (SQL): de facto query language and data access standard supported by the majority of DBMS vendors
- Database communication interfaces
  - Accept end-user requests via multiple, different network environments

Managing the Database System: A Shift in Focus

- Disadvantages of database systems
  - Increased costs
  - Management complexity
  - Maintaining currency
  - Vendor dependence
  - Frequent upgrade/replacement cycles

Preparing for Your Database Professional Career

TABLE 1.3	DATABASE CAREER OPPORTUNITIES	
JOB TITLE	DESCRIPTION	SAMPLE SKILLS REQUIRED
Database Developer	Create and maintain database-based applications	Programming, database fundamentals, SQL
Database Designer	Design and maintain databases	Systems design, database design, SQL
Database Administrator	Manage and maintain DBMS and databases	Database fundamentals, SQL, vendor courses
Database Analyst	Develop databases for decision support reporting	QL, query optimization, data warehouses
Database Architect	Design and implementation of database environments (conceptual, logical, and physical)	DBMS fundamentals, data modeling, SQL, hardware knowledge, etc.
Database Consultant	Help companies leverage database technologies to improve business processes and achieve specific goals	Database fundamentals, data modeling, database design, SQL, DBMS, hardware, vendor-specific technologies, etc.
Database Security Officer	Implement security policies for data administration	DBMS fundamentals, database administration, SQL, data security technologies, etc.
Cloud Computing Data Architect	Design and implement the infrastructure for next-generation cloud database systems	Internet technologies, cloud storage technologies, data security, performance tuning, large databases, etc.
Data Scientist	Analyze large amounts of varied data to generate insights, relationships, and predictable behaviors	Data analysis, statistics, advanced mathematics, SQL, programming, data mining, machine learning, data visualization

Summary

- Data consists of raw facts and is usually stored in a database
  - Database design defines the database structure
    - Can be classified according to the number of users, location, as well as data usage and structure
  - Databases evolved from manual and computerized file systems
    - There are some limitations of file system data management
    - DBMSs were developed to address the file system's inherent weaknesses

HW 1

- Chapter 1 Application
  - Due Aug 29 11:59pm – see Canvas
- Chapter 1 Class Exercises due Wednesday
- **Quiz 1 first 15 minutes on Wednesday – BRING LAPTOP**
  - Practice Quiz